(2)

RECOGNIZING, THINKING AND LEARNING AS INFORMATION PROCESSES

Technical Report AIP - 99

Herbert A. Simon & Qicheng Jing

Department of Psychology
Carnegie Mellon University
Pittsburgh, PA 15213

The Artificial Intelligence and Psychology Project

Departments of
Computer Science and Psychology
Carnegie Mellon University

Learning Research and Development Center
University of Pittsburgh

RECOGNIZING, THINKING AND LEARNING AS INFORMATION PROCESSES

Technical Report AIP - 99

Herbert A. Simon & Qicheng Jing

Department of Psychology
Carnegie Mellon University
Pittsburgh, PA 15213

30 August 1987

This research was supported by the Computer Sciences Division, Office of Naval Research, under contract number N00014-86-K-0678. Reproduction in whole or part is permitted for any purpose of the United States Government. Approved for public release; distribution unlimited.

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; Distribution unlimited
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) AIP - 99	5 MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Carnegie-Mellon University / 6b OFFICE SYMBOL (If applicable)	7a NAME OF MONITORING ORGANIZATION Computer Sciences Division Office of Naval Research
6c. ADDRESS (City, State, and ZIP Code) Department of Psychology Pittsburgh, Pennsylvania 15213	7b ADDRESS (City, State, and ZIP Code) 800 N. Quincy Street Arlington, Virginia 22217-5000
8a. NAME OF FUNDING / SPONSORING ORGANIZATION Same as Monitoring Organization / 8b. OFFICE SYMBOL (If applicable)	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-86-K-0678
8c. ADDRESS (City, State, and ZIP Code)	10 SOURCE OF FUNDING NUMBERS p4000ub201/7-4-86 PROGRAM ELEMENT NO: N/A; PROJECT NO: N/A; TASK NO.: N/A; WORK UNIT ACCESSION NO: N/A

11. TITLE (Include Security Classification)
Recognizing, Thinking, and Learning as Information Processes

12 PERSONAL AUTHOR(S)
Herbert A. Simon & Qicheng Jing

13a. TYPE OF REPORT	13b TIME COVERED	14 DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Technical	FROM 86Sept15 TO 91Sept14	1987 August 30	22

16 SUPPLEMENTARY NOTATION

17 COSATI CODES			18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	physical symbol systems, perception, thinking, learning information processing

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

SEE REVERSE SIDE

20 DISTRIBUTION / AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☐ DTIC USERS	21 ABSTRACT SECURITY CLASSIFICATION
22a NAME OF RESPONSIBLE INDIVIDUAL Dr. Alan L. Meyrowitz	22b TELEPHONE (Include Area Code) (202) 696-4302 / 22c. OFFICE SYMBOL N00014

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

Unclassified

ABSTRACT

The physical symbol system hypothesis explains how cognitive processes can be realized by a physical system. Recognition, representing pictorial information, thinking, and even processes like intuition, discovery, and learning can be carried out both by human and by articfical symbol systems. The symbol system hypothesis can be extended to explain social interactions of individuals in society, for when enough cultural-historical knowledge is stored in such a system, it should be able to respond to the social situation.

ABSTRACT

The physical symbol system hypothesis explains how cognitive processes can be realized by a physical system. Recognition, representing pictorial information, thinking, and even processes like intuition, discovery, and learning can be carried out both by human and by artificial symbol systems. The symbol system hypothesis can be extended to explain social interactions of individuals in society, for when enough cultural-historical knowledge is stored in such a system, it should be able to respond to the social situation.

Recognizing, Thinking, and Learning as Information Processes

Herbert A. Simon
Carnegie-Mellon University
and Qicheng Jing
Chinese Academy of Sciences

In the history of psychology, significant strides of progress have often been made by incorporating theories and concepts borrowed from other branches of natural science. This has been particularly true since debates of "schools" of psychology have subsided and research has turned to the accumulation of facts, and has taken a more interdisciplinary approach toward strengthening psychology's scientific character (Jing, 1962, 1982).

During the past 30 years, great progress has been made in accounting for cognitive processes as processes carried out by a physical symbol system. These developments, which have given rise to the discipline now usually called cognitive science, provide an explanation of what goes on in a human being (or a computer) in the course of recognizing an object or a word, solving a problem, or acquiring new knowledge or skills. Cognitive science has philosophical implications for resolving the epistemological question of how knowledge is acquired by the human mind and for providing a better treatment of the mind-body problem. Cognitive science draws upon, and contributes to, cognitive psychology, artificial intelligence, linguistics, and even some parts of anthropology and sociology.

In this paper we shall first describe symbol systems and lay out the general principles of the physical symbol system hypothesis, and the philosophy underlying it.

Second, we shall discuss the basic processes whereby a physical symbol system recognizes stimuli, and stores knowledge in association with the symbolic labels attached to the recognized stimuli. The description of these processes will permit us to consider more closely the nature of symbols in cognition. Moreover, we will see that these recognition processes provide an explanation for the kind of thinking we call intuitive.

Next, we shall describe the other processes of thinking and reasoning, distinguishing them from the processes of formal logic. This discussion will allow us to consider, also, the nature of non-standard logics: for example, modal logics, non-monotonic logics, and dialectical reasoning.

Finally, we shall describe some of the kinds of adaptive processes that symbols systems can (and do) employ for learning and discovery.

Physical Symbol Systems

In the information processing approach to cognition, it is required that all processes postulated to occur in the course of recognizing, thinking, and learning be implementable by a physical symbol system. But the *way* in which they are implemented is left completely open; the physical symbol system may be a human brain or its analogous counterpart, a computer, and if it is a computer, its basic memory elements may be magnetic cores, chips, or vacuum tubes. What is essential is only that the system be capable of manipulating *patterns* of all sorts -- of inputting ("reading") patterns, of storing patterns in a memory, of building complex patterns from simpler components, or outputting ("writing") patterns, of comparing patterns to see if they are the same or different, and of conditioning its behavior ("branching") upon the outcomes of such comparisons.

Patterns can therefore be represented by any manipulable arrangements of matter. What identifies them is the arrangement, not the special matter of which they are composed. For example, the pattern called "torus" is exemplified by any hollow ring or cylinder, whether it be made of iron, wood, plastic, or drawn in a picture.

When patterns denote (point to, represent) other patterns or external objects or relations among objects, they are called *symbols*. Thus, some of the patterns within the physical symbol system are symbols that represent or reflect external reality. The idea of symbols is important in explaining how the mind provides a subjective reflection of objective reality. As Marx puts it, "ideological things are nothing other than material things introduced into the human brain that have been transformed in the human brain (Marx, 1972)."

Thus, the physical symbol system hypothesis supports materialistic monism in asserting that things in the objective world are primary and mental events secondary, the former being the source of knowledge and the latter the subjective reflection of external reality. Mind is the product of matter. Compare Lenin (1952): "Thought is a function of the brain. Sensations, i.e., the images of the *external world*, exist *within us*, produced by the action of things on our sense organs. . . the mind does not exist independently of the body. Mind is secondary, a function of the brain, a reflection of the external world." Similarly, in the case of the computer, the physical symbol system hypothesis asserts that the computational capabilities provided by symbolic processes rest on the mechanisms of the computer hardware.

The Physical Symbol System Hypothesis

The *physical symbol system hypothesis* is the hypothesis that the necessary and sufficient condition that a system be capable of exhibiting intelligence is that it have the symbol manipulating capabilities just enumerated. The hypothesis has two important consequences. If the hypothesis is correct, then computers have the capability of exhibiting intelligence. And if the hypothesis is correct, then the brain is (at least) a physical symbol system.

Mind and Body. One interesting and important feature of the hypothesis is that, while it postulates a physical mechanism (brain or computer) as essential for the manipulation of symbols, it specifies only abstract properties of this mechanism. Thus, the level of symbols can be distinguished from the underlying level of physical mechanism and its material realization, and the hypothesis thereby provides a solution to the mind-body problem that removes the mystery of how thought can be produced by the physical matter of the brain. This solution avoids the criticisms that may be leveled against idealism as well as those that may be leveled against "vulgar" materialism.

In contrast to idealism, the interpretation of the mind-body problem provided by symbol systems permits the external world to be represented (reflected) in mind as a consequence of the mind's (i.e., the physical symbol system's) capability for receiving sensory inputs, together with its capability for building symbol structures in memory to denote objects and relations in the external world.

The symbol system interpretation proposed here, in contrast to "vulgar" materialism, distinguishes the level of symbols and symbol structures, realizable by a considerable range of physical systems, from the underlying material in which they are realized. Thus, the same concept, abstract or concrete, can be represented in the brain by a pattern of neuronal activity, in a computer of the 1950's by a pattern of excitations in a bank of vacuum tubes, or in a contemporary computer by a bit pattern in a chip. Since it is the pattern, not the material, that defines a symbol, these totally different physical devices can represent identical symbols and symbol structures, thereby giving a precise characterization of mind (concepts embodied in symbol structures) as a product of matter (the physical devices that realize the symbols), but distinct in kind from the matter that realizes it.

These symbols of a physical symbol system may represent different levels of abstraction. They may be inexact copies of external reality: perceptions, imagery, representations of objects. They may also be more generalized representations: e.g., concepts of things and events. Creating and operating upon the latter requires more elaborate manipulation of information stored in memory than is required at lower levels of abstraction. I. P. Pavlov designated the more abstract level that handles natural language as the second signal system. A word is a higher level signal denoting direct perception of an object. The first signal system receives sensory stimuli that are of biological importance for the survival of the organism. With the evolution of the human species, the second signal system evolves on the basis of the first signal system: it is the signal of signals (Pavlov, 1951(a); Ivanov-Smolenski, 1963). The second signal system represents external reality at a higher level of generalization and abstraction than the first. In cognitive psychology, knowledge related to what Pavlov has termed the second signal system is a major focus of research.

Finally, the symbol system interpretation allows no dualism, in which mind and matter pursue their separate paths; for mind, in the form of symbol structures, is always realized by a physical symbol system, and cannot exist independently of such a realization. The relation between a physical symbol system and symbol structures is a relation of cause and effect.

Thinking and Logic. Another important feature of the physical symbol system hypothesis is that it provides an interpretation of human thinking and reasoning that shows reasoning to be related to, but much more flexible and general than the processes of formal logic. Formal logic considers only symbols that represent sentences and denote propositions. It excludes representations of reality of the sort we commonly call "imagery." Formal logic specifies a small set of fixed inference rules. Only those additions to or changes in the set of sentences are admitted that can be produced by an application of an inference rule. Hence, formal logic excludes the kinds of changes in images that correspond to physical or other processes acting upon the object that is imaged. It cannot model directly a system's development through time. By restricting thinking to steps produced by the inference rules, it also bars those sudden leaps of thought that we generally call "intuition" or "inspiration." For all of these reasons, the processes of thinking are far more extensive than those of logic.

As we shall see, physical symbol systems are not subject to any of these limits. They can represent reality in a variety of modes, including imagery. Arbitrarily rich sets of operators can be introduced to change images in ways that correspond to the operation of natural laws upon the systems being imaged. The operators are not tautological, like the inference rules of logic, but may incorporate any kinds of empirical laws as well as logical principles. And because a single operator can incorporate arbitrary amounts of knowledge, its application to a situation may produce an intuitive leap of thought. Thus, in addition to reasoning in the ways usually considered "logical," physical symbol systems may employ imagery and may proceed by intuition (Larkin and Simon, 1987).

Other Symbol Systems: DNA. The computer and the brain are not the only physical systems that are recognized today to be symbol systems. A third exmaple and a very instructive one, is the chromosomal material, DNA, which contains the information used by the cell to synthesize proteins. The information in DNA reflects the real structure and functioning of the organism (with remarkable accuracy). Moreover, it is the pattern of nucleotide sequences, not the internal chemistry, that conveys the information. The information is transcribed into similar sequences of RNA, where the pattern serves as a template for laying down the sequences of amino acids that make up the protein molecule.

Of course, the physical symbol system that we call the chromosome has very different functi... hence different capabilities. from the brain or the computer. Its functions are. first. to reflect the structure of protein molecules in order to instruct the process of protein manufacture. and. second. to permit its own replication. Nevertheless. it illustrates admirably how pattern. realized in particular material. the DNA. can serve to hold information about a completely different material. the protein. It provides a concrete model that shows how symbols can be produced by matter although the two are not identical. The symbol is not the physical substrate: it is the pattern in which that substrate is arranged.

Recognition Processes

The linkage of mind to the outside world (reflection) is provided by sensory channels that lead to *recognition* of familiar objects and relations. In this way, information about external reality can be used to build up models of that reality in memory, and to use new information to modify and augment models already stored in memory by past experience.

It is essential that the recognition system contain some active mechanism. for the reflection process is not passive (Lomov, 1984). The system may start with a directional response to a stimulus. of the nature of the so-called "orienting reflex" (Pavlov, 1951(b); Sokolov, 1963). If the stimulus catches the interest. or corresponds with the preset scanning objective of the system. then it extracts *features* from the stimuli that impinge upon the sense organs (McCulloch. Lettvin. Maturama and Pitts, 1959: Hubel and Wiesel. 1959). By some process (as we shall see, several candidate processes have been proposed), these features are used to *identify* the stimulus if it belongs to some already familiar object or system of objects, that is, to relate it to the familiar object or objects. If it is not already familiar, the features can be stored as a new class of objects (concept learning).

Models of the Recognition Mechanism

One theory of pattern identification (Selfridge, 1959) hypothesizes that the stimulus features are compared in parallel with all stored patterns, and the pattern that matches the stimulus features most closely is selected. This DEMON theory attaches weights to the features of each identifiable class of stimuli and cross-correlates the features of the incoming stimulus with these weights, thus allowing partial matching.

Some contemporary descendants of the DEMON model represent memory as a network of nodes and links, a variable strength being associated with each link. Stimulus features and identifiable concepts are represented by nodes, and a stimulus is recognized as belonging to a particular concept if, in aggregate, it is more closely linked to that concept than to others. This CONNEXIONIST model (Rummelhart and McClelland, 1986), like the DEMON model, assumes massive parallel, simultaneous, action of the entire system.

A somewhat different theory (EPAM), postulating serial rather than parallel action, arranges tests for the presence or absence of stimulus features in a discrimination net -- a kind of "twenty questions" sorting system (although the branches need not be binary). Stimulus objects are sorted through the net to find a concept that (partially) matches them (Feigenbaum and Simon, 1984.)

There is today no conclusive experimental evidence to show which one of these theories, or some other, provides the true explanation for recognition. All the theories, whether serial or parallel, presuppose a network in the physical symbol system that enables it to find correspondences between stored knowledge of what "things" should look like and the visual information presented to the sense organs. The system need not code the absolute position of features, but perhaps relations of stimulus features to a consistent perceptual reference frame, such as the orientation of the axis of a shape (Humphreys, 1987). Even if the same object is viewed twice, it may not be observed from the same viewpoint and its shape may thereby be distorted. Any recognition system needs to abstract the essential information defining the object it is viewing.

The fact that all three theories described above have been programmed successfully for computers demonstrates that recognition of objects, relations among objects, and systems of objects can be achieved by a physical symbol system. Moreover, more detailed examination of these realizations shows that the recognition times they would achieve, if the speeds of their component processes were comparable to those in the human nervous system, are similar to those exhibited by human subjects -- that is, they are in the neighborhood of a half second. Indeed, most current neurophysiological research clearly favors network models over a mass-action approach. Research results also show that the vertebrate brain possesses enormous plasticity and adaptive capacity, so that storage of memories may be mediated by structural as well as functional changes in neurons (Rosenzweig, 1985).

Expertise and Intuition

The study of human expertise in such domains as chessplaying, solving physics problems, and making medical diagnoses (De Groot, 1965; Chase and Simon, 1973; Simon and Simon, 1978; Elstein, Shulman, and Sprafka, 1978) has demonstrated that the ability frequently exhibited by experts of making correct judgments intuitively (that is, rapidly and without deliberate calculation) can be explained fully by any of these theories of recognition.

More precisely, it has been shown (e.g., Chase and Simon) that, in the process of acquiring high skill, the first-class expert (chess master or grandmaster, physician) learns at least 50,000 patterns in the domain of expertise that can be recognized on sight, and that also allow the expert to access and recall from long-term memory a larger or smaller body of information that has been learned and associated with the patterns.

Thus, a skilled diagnostician, confronted with a patient who recounts symptoms or exhibits them, will recognize in the symptoms a familiar pattern, and will thereby recall from memory one or more hypotheses about the diseases that might be producing the symptoms. In addition, the physician will recall what additional tests should be performed to make the diagnosis definitive, and what course of treatment should be followed when the diagnosis has been made. And all of this can be done in a few seconds -- that is, intuitively.

We cannot review here the considerable body of empirical evidence that shows that intuition is nothing more (and nothing less) than the exercise of the recognition capability possessed by any trained mind -- any mind that has acquired its 50,000 patterns and associated knowledge schemas. A review of some of the evidence can be found in Simon, 1986.

Sometimes acts of human intuition are so impressive that we apply to them terms like "creative" or "inspired." Today there is no reason to think that creativity or inspiration require different mechanisms from those used in ordinary problem solving. On the contrary, there are two substantial bodies of evidence for the thesis that the mechanisms are the same. First, there are now some examples, in the domains of drawing and music, of computer programs that create works of art whose esthetic qualities can be appreciated by human beings. In drawing, I refer especially to the program called AARON, the work of the painter, Harold Cohen, which makes drawings that create the illusion of space, of abstract objects, of human figures, of plants, all arranged in esthetically pleasing ways that are strongly suggestive of social interaction. In music, I can mention very early work, dating to the 1950's, of the program on the ILLIAC computer, that composed the ILLIAC SUITE and, of more esthetic interest, the COMPUTER CANTATA. The programs that do this are comparable to other "expert systems" that have been constructed to do more prosaic tasks.

The second line of evidence are programs that have capabilities for discovery in mathematics and science. Among these, we may mention Meta-DENDRAL, which finds low-level laws of mass spectrography, MOLGEN, which designs experiments in molecular genetics, AM and EURISKO, which find new mathematical concepts, BACON, GLAUBER, STAHL, and DALTON, which induce laws from empirical data, and KEKADA, which plans strategies of experimentation (See Langley, Simon, Bradshaw, and Zytkow, 1987).

To deny that these programs exhibit creativity, or even inspiration, we would have to deny these same qualities to people when they produce similar works of art or science. Of course, the programs use several modes of thinking, in addition to those we have been describing as intuitive. We turn now to a more general discussion of what these various kinds of thinking processes are.

Thinking and Reasoning

Historically, thinking, especially "logical" and mathematical thinking, has been associated with formal logic and the rigor of mathematics. This view represents such an incomplete truth as to be no truth at all. The error arises from confusing the formal inference rules of logic and mathematics, whose proper function is to test the validity of conclusions already reached, with the processes used to *discover* new truths that can be inferred from those already held.

We have already discussed one kind of thinking -- intuitive thinking -- that lies outside the realm of logic, as usually conceived. In this section, we will introduce other thinking processes, including heuristic search, the use of imagery and non-standard forms of logic. In the course of the discussion, we will also show that computers are not limited to logical reasoning, and in fact are capable of doing all of the kinds of thinking we describe. Hence they are available for the simulation of human thinking in all of its forms.

Logic versus Heuristic Search

Most of the everyday activities, including professional activities, that we call thinking or reasoning are concerned with discovering what is true, and only a small fraction of these activities with verifying the discoveries. Thus, when we examine the thinking-aloud protocols of subjects who are solving difficult problems, we find that most of their activity can be described and explained as a search, almost always highly selective, through a large (sometimes enormous) space of possibilities. Instead of the small number of inference rules that are admitted in most systems of formal logic, problem solvers employ a large repertory of inference rules that incorporate not only accepted principles of logic but a great deal of knowledge of the subject-matter domain, as well.

As a consequence, the numerous tiny steps that are required for a formal proof are replaced, in real human reasoning, by a much smaller number of macrosteps, each involving a possibly complex inference that may be valid only in a restricted task domain. If the macro-operators that take these steps are sufficiently powerful, as they usually are when the problem solver is an expert, then an impossibly exhausting search in the problem space is replaced by a very limited search, or even by a systematic movement to the goal, with no trial-and-error steps at all.

Parenthetically, we observe that programming languages like PROLOG, which are built on the metaphor of thinking as logical derivation, provide less satisfactory descriptions of human problem solving than do list processing languages, like LISP, or production system languages, like OPS5. These latter languages are closer to human thinking because they permit -- even encourage -- the introduction of arbitrarily large sets of powerful, taskdependent operators for searching the problem space.

Nor need the operators used for searching a problem space always be infallible. Instead, they may be *heuristic* (sometimes also called "intuitive"), usually leading to valid conclusions, but not always. The use of fallible heuristics accounts for the fact that after finding problem "solutions," people often test these solutions in various ways for validity. If only rigorously justified inference rules were used in reasoning, then incorrect conclusions would have to be attributed solely to "bugs" in the machinery of application. But studies of human errors (Brown and Van Lehn, 1980) show them to be much richer in variety and origin than this view admits. The fallibility of heuristics provides a partial explanation of the ubiquitous need to check conclusions reached by human reasoning.

Non-Standard Logics

It is sometimes proposed that the inadequacies of standard formal logic for describing human reasoning should be overcome by introducing modified non-standard logics. Various candidates have been proposed: "fuzzy" logics (Zadeh, 1975), modal logics (McCarthy and Hayes, 1969), non-monotonic logics (Doyle, 1979; Weyhrauch, 1980), and others. But there is a much simpler and far more adequate solution. If we simply view the set of heuristic operators used for searching a task domain as a "logic" for that task domain, we solve the problem at one stroke. The problem is not to define new formal logics, but simply to represent directly the powerful information processes, mentioned in the previous section, that human beings normally adopt when they are reasoning or solving problems.

Are Computers Inherently "Logical"?

This solution to the problem of explaining reasoning -- treating it as heuristic search -- not only fits the empirical evidence, but also brings with it a number of valuable bonuses. First, it gets rid of the fallacy of supposing that any reasoning or thinking that can be done by a computer is inherently and necessarily "logical," and hence that computers are incapable of simulating processes of human thinking that are not constricted by the inference rules of any formal logic. On the contrary, the operators employed for heuristic search can depart as far as we please from the forms of formal logic, even to the point of simulating all kinds of "illogical" and fallacious thinking. There is no difficulty, for example, in programming a chessplaying program or a medical diagnosis program to jump to conclusions on the basis of insufficient evidence -- either for no reason at all or because its computational time is limited.

Dialectical Reasoning

A second bonus from using heuristic search as our general model of human thinking is that it enables us to provide unequivocal interpretations of difficult and complex reasoning processes. If we wish, for example, to give a program the ability to use dialectical methods in its thinking, we must arrive at a definition for these methods. By way of illustration, let us suggest several, possibly complementary, representations of dialectical reasoning.

We will use the language of production systems. That is, we consider associations between sets of conditions, C_i ($i = 1,\ldots,n$), such that action A_i is taken whenever conditions C_i are satisfied. We then write $C_i \dashrightarrow A_i$. Now we may take as a thesis, $C_1 \dashrightarrow A_1$, and as an antithesis, $C_1 \dashrightarrow A_2$, where $A_1 \neq A_2$. Here, satisfaction of C_1 brings about both the action A_1 and the antithetical action A_2.

One way in which a synthesis can be built on these two antithetical productions is to introduce new conditions to discriminate between the circumstances under which A_1 should occur and those under which A_2 should occur. For example, we could build $C'_1 = C_1 \wedge D \dashrightarrow A_1$, and $C''_1 = C_1 \wedge \neg D \dashrightarrow A_2$.

Another possibility for dialectical method arises when an action is taken to achieve certain results, $A \supset R$, the thesis, but it appears that the action will also bring about, as an unintended or undesired consequence, a further result, $A \supset U$, the antithesis. Now a synthesis could be achieved if we could construct a modified action, A', such that A' would produce R but not U: $A' \supset (R \wedge \neg U)$.

A slight modification of this procedure would cause the unintended consequence itself to trigger the synthesis. That is, we assume that A implies both R and U; while U implies B, which replaces A. Here $A \supset R$ is the thesis; $A \supset U$ is the antithesis; and $U \supset B$ is the synthesis. Such a procedure could serve as a formalization for theories that involve contradictions and dialectical processes in the development of sequences of "stages" in a system.

These are just three possible interpretations of dialectical reasoning that would permit the statement of a thesis and an antithesis, and the resolution of the contradiction through synthesis. Other dialectic-like processes can be found in the methods for resolving contradictions that are used in so-called non-monotonic logics (Doyle, 1979; Weyhrauch, 1980).

Pictorial Thinking and Imagery

A third byproduct of the method of heuristic search is that it gives us a way of introducing pictorial representations and diagrams, and showing how these are used in thinking and reasoning, and how they may facilitate intelligent perception and the solving of problems (Simon and Barenfeld, 1969; Larkin and Simon, 1987). Pictures and diagrams have both an internal and an external aspect, for people use diagrams in books or on paper to assist them in solving problems, but they also form mental images for the same purpose.

While the amount of information that can be held in a single mental image is rather limited compared with the amount that can be recorded in a diagram on paper, the kinds of information that can be obtained from image and diagram are quite similar, as is the facility with which various kinds of inferences can be drawn from them. For this reason, after some preliminary comments about the representation of mental images in a physical symbol system, we will discuss images and diagrams together, without distinguishing between them.

Representation of Mental Images. There are a number of ways in which mental images can be represented in a symbol system. Kosslyn (1980) has discussed these alternatives rather fully, and has shown that the human mind probably uses at least two such representations. One is a raster of discrete points that operates as a "photograph" of a diagram or picture, and from which symbolic processes can extract features. This raster might be thought of as a copy of the pattern of excitations on the retina, but the image can be created by information drawn from memory as well as by information in an external display.

A second mode of representation of images employs symbol structures in a slightly more abstract way. These structures consist of nodes and of links that connect these nodes (list structures), joining them in such a way that the nodes correspond to elements in a diagram or picture, while the links correspond to the spatial relations among these elements. For example, a mental picture of a system of pulleys holding up some weights would include a symbol to represent each weight, each pulley, and each major segment of the ropes from which the weights and pulleys were suspended. Then each pair of elements that are directly connected and that exert forces on each other would be joined in the mental picture by a link.

The two forms of imagery, raster and symbol structure, can be supposed to operate conjointly as the "mind's eye."

Uses of Imagery. The mind's eye and external diagrams contribute powerfully to thinking and reasoning in two ways (Larkin and Simon, 1987). First, the very process of creating the mental picture or external diagram makes explicit many objects and relations that may be only implicit in the information from which the picture is formed.

If we are told to consider the two diagonals of a rectangle, we not only envisage the perimeter of the rectangle and the diagonals, but also the point of intersection of the diagonals. This point of intersection was not mentioned in the verbal description of the diagram, but it springs instantly (and effortlessly) to view when we image the rectangle, just as it does when we draw the rectangle and diagonals on paper. In both cases, internal and external, the process of imaging provides new information (makes new inferences) in a computationally efficient way.

But this is not all. Once drawn or imaged, the diagram provides means for making further inferences with far less computation than would be required to make them in propositional form. This is accomplished through the kinds of recognition processes that have already been discussed.

Knowledge of the task domain allows certain features of the diagram to be recognized, and information associated in memory with these kinds of features allows appropriate inferences to be made. To take a simple example, suppose two weights hang from the two ends of the rope over a pulley. A person with an elementary knowledge of physics who images this situation will notice the connection of the first weight with one end of the rope, and will thereby "see" (i.e., infer immediately by recognition) that the force exerted by the rope must equal the weight. Similarly, he or she will "see" that the forces on the two halves of the rope are equal.

The mind's eye has yet another advantage over external diagrams in that mental images can be manipulated at will to assist in planning the arrangement of objects in space, for example, alternative ways of arranging furniture in an empty room. The pieces of furniture may be imaged in various locations, paying due regard to their spatial dimensions and esthetic properties. Through just this kind of manipulation of images of parts of machines or instrument, scientists and engineers sometimes solve technical problems and even make inventions.

It is important to note that the "seeing" involves inference not only of spatial relations, but also of physical relations: the recognition operators incorporate physical laws as well as geometric ones. Obviously, what properties of the system will be recognized depends on what knowledge the imager has of the physical laws of the situation.

Summary: The Forms of Thinking

In sum, the physical symbol system hypothesis and the concept of heuristic search provide us with a realistic description and explanation of thinking. This explanation shows thinking to be quite different, ordinarily, from the deductive processes of logic, and readily accommodates a wide variety of modes of thinking, which a human being can use separately or conjointly. One of these modes is intuitive thinking, which makes use of recognition processes, hence is based upon previously acquired knowledge.

Another mode of thinking is heuristic search, which can make use of the recognition processes in the selection of search operators, and also makes use of general heuristics like means-ends analysis to guide its search. Thinking we call logical or analytical usually conforms to this model of heuristic search. The inference operators that are employed may go far beyond standard logic, or even non-standard logics. In fact, they are generally not tautological, but incorporate knowledge of the subject matter that is being reasoned about.

Still another mode of reasoning, usually in conjunction with heuristic search, makes use of mental images to model external reality, often with the additional help of external diagrams or pictures.

These modes -- intuitive thinking, heuristic search, and thinking with imagery -- may not exhaust all of these processes that we call "thinking" in human beings. But simulations incorporating these modes of thinking have been shown to be capable of encompassing a wide range of tasks to which human beings apply their minds.

Learning and Doing

A theory of human cognition must explain not only how people perform difficult thinking and problem-solving tasks, but also how they learn to perform them. During the first decades of cognitive science research, the main emphasis was placed on understanding performance, and only recently has attention turned in a major way to the processes whereby knowledge and skills are acquired, or to the processes of discovering new knowledge. Current information processing research on machine learning, some of it primarily concerned with artificial intelligence, but some directed toward human learning, can be found in the published proceedings of two *Machine Learning* workshops, in the new journal, *Machine Learning*, and in the journal, *Cognitive Science*.

If by learning we mean any change in a symbol processing system that improves its performance on some class of tasks, then there can be many kinds of learning, for there are many ways in which a complex system can be improved. A mere addition of new information in literal form to the knowledge base, without any significant alteration in the information, is usually called *rote learning*. We are less interested in rote learning than in learning that provides some genuine understanding of the material learned.

The forms of learning by understanding that have been studied in recent years bear a close relation with problem solving and discovery, so that we can discuss learning in terms of the latter concepts. We will illustrate this point by describing two special, but important, forms of learning: *learning from examples*, and *learning by doing*.

Suppose that we are supplied with an example, worked out step by step and finally solved, of a problem of a certain kind. Then, if by examining the example, we can learn to solve s...ilar problems, we call this learning by example. If, on the other hand, we are simply given the problem and allowed to struggle with it until we solve it, and then are able to solve other similar problems, we call this learning by doing. It is obvious that learning by doing is a discovery process that depends on problem solving processes.

Learning by doing is closely connected with learning by example; for when we have succeeded in solving a problem, the solution itself -- stripped of the unnecessary false steps we took along the way -- is now a worked-out example, and can be used to aid learning just as an example provided by the textbook or the teacher can.

Now let us see how learning by example works. We consider the much-studied case of solving a single linear algebraic equation, using the following example:

7X + 5 = 3X + 13	If N on left	-->	Sub(N)
7X = 3X + 8	If NX on right	-->	Sub(NX)
4X = 8	If NX on left, $N \neq 1$	-->	Div(N)
X = 2	If "X=N"	-->	Halt & Check

On the left-hand side, we see the solution worked out, step by step. On the right-hand side, we see four productions (condition-action pairs) that show what actions were taken and what cues triggered them. The answer is an expression of the form, "X equals a number." When such an expression is reached, the solver halts and checks the solution by substituting it back in the original equation. (See the fourth production or the last line.) The original expression is not in that form; among other things, it has a number on the left-hand side. The first production notices this ("N on left") and subtracts that number from both sides of the equation ("Sub(N)"). Now the equation still has an expression of the form "NX" on its right-hand side, something that should be absent from the solution. The second production notices this ("NX on right") and subtracts 3X from both sides ("Sub(NX)"). Now the equation differs from the desired form only in having a coefficient of 4 instead of 1 for X. The third production notices this ("NX on left, $N \neq 1$") and divides both sides by 4.

We would postulate, and evidence supports the postulate, that when a student has learned to solve equations of this sort, he or she has stored in memory a set of productions like those shown above. This does *not* mean that the student has memorized the rules that express the productions in natural language. It means that the student has learned to notice the presence of the cues (conditions) on the left-hand sides of the rules, and has learned to take the appropriate actions, on the right-hand sides, whenever the corresponding cues are noticed. What has been learned is a set of actions that are triggered by perceptual recognition of cues -- a means of solving the equations "intuitively," or with "insight."

But how can the productions be acquired from the example? A student who looks at a pair of successive lines of the example can discover what change has been effected from one line to the next -- the disappearance of the constant from the left side, for example, or of the coefficient of X. The "reason" for this change can be induced by noting that the transformed equation more closely resembles the desired result, "X = N," than did the previous one. The action that is associated with the change -- an action known from previous study not to change the value of X in the equation -- clearly effected the removal of this difference between current and desired equation. Thus, by application of means-ends analysis, the example can be understood and the appropriate productions induced from it.

David Neves (1958) wrote a computer program that simulated this learning. When shown the example, the program proceeded to examine it, noticing the cues and actions and constructing the corresponding productions. Having added these productions to its memory, it then could solve equations of the same general form. If this is indeed a viable learning procedure for people, then it should be possible to construct a curriculum for learning by example, in which students are simply presented with a carefully designed sequence of worked-out examples and problems, and allowed to develop their algebra skills from them. Such a curriculum has indeed been worked out under the direction of scientists of the Institute of Psychology, Chinese Academy of Sciences, and in initial trials appears to support very efficient learning of middle-school students.

Conclusion

The physical symbol system hypothesis enables us to explain how cognitive processes, human or otherwise, can be realized by a physical system. The world is material; the human brain and the computer are both material symbol systems, differing only in the degree of sophistication of their symbol structures. The human brain can generate thought, as its function or product. Similarly the computer can generate intellectual processes. In this paper we have tried to show how recognition, representing pictorial information, thinking, and even sophisticated processes like intuition, discovery, and learning can be carried out by human or artificial symbol systems.

In the future the physical symbol system hypothesis may be extended to explain social interactions of individuals in society. Since a physical symbol system stores knowledge and, through heuristic search, takes appropriate actions in response to stimuli, it follows that, when enough cultural-historical knowledge, including knowledge of the social and interpersonal situation, is stored in a physical symbol system, it should be able to decide and act in ways that respond to the social situation.

All materialists, including dialectical materialists, insist that physical processes underlie human conscious activities. There is no reason to believe that it will be impossible, for all time, to build a computerized physical symbol system that behaves as though conscious: a computer control system that incorporates motivation, emotions, and intentionality, and that directs short-term activities and plans long-term activities. (See Burks, 1987.)

References

Brown, J. S., and K. VanLehn. 1980. Repair theory: A generative theory of bugs in procedural skills. *Cognitive Science*, 4:379-426.

Burks, A. W. 1987. Teleology, intentionality, and consciousness. *Acta Psychologica Sinica*, 19:156-166 (In Chinese).

Chase, W. G., and H. A. Simon. 1973. Perception in chess. *Cognitive Psychology*, 4:55-81.

De Groot, A. D. 1965. *Thought and Choice in Chess*. The Hague: Mouton.

Doyle, J. 1979. A truth maintenance system. *Artificial Intelligence*, 12:231-272.

Elstein, A. S., L. S. Shulman, and S. A. Sprafka. 1978. *Medical Problem Solving*. Cambridge, Mass.: Harvard University Press.

Feigenbaum, E. A., and H. A. Simon. 1984. EPAM-like models of recognition and learning. *Cognitive Science*, 8:305-336.

Hubel, D. H., and T. N. Wiesel. 1959. Receptive fields of single neurons in the cat's striate cortext. *J. Physiol.*, 148:579-596.

Humphreys, G. W. 1987. Objects, words, brains and computers: Framing the correspondence problem in object and word recognition. *Bull. Brit. Psychol. Soc.*, 40:207-210.

Ivanov-Smolenski, A. G. 1963. *Experiments in the Objective Study of Functions and Interactions of Signal Systems of the Brain*. Moscow: Government Medical Literature Publishers (In Russian).

Jing, Q. 1962. Natural science and psychological theory. *New Construction*, No. 12, 31-46 (In Chinese).

Jing, Q. 1982. Perspectives in the development of psychology. *Psychological Information*, No. 4, 7-10 (In Chinese).

Kosslyn, S. M. 1980. *Image and Mind*. Cambridge, MA: Harvard University Press.

Langley, P., H. A. Simon, G. L. Bradshaw, and J. M. Zytkow. 1987. *Scientific Discovery*. Cambridge, MA: The MIT Press.

Larkin, J., and H. A. Simon. 1987. Why a diagram is (sometimes) worth 10,000 words. *Cognitive Science*, 11:65-100.

Lenin, V. I. 1952. *Materialism and Empirio-Criticism*. Moscow:

Foreign Languages Publishing House. 85.

Lettvin, J. Y., H. R. Maturana, W. S. McCulloch, and W. H. Pitts. 1959. What the frog's eye tells the frog's brain. *Proc. IRE*, 47:1940-1951.

Lomov, B. F. 1984. *Methodological and Theoretical Problems in Psychology*. Moscow: "Science" Publishers. 131-132 (In Russian).

Marx, K. 1972. *Selected Works of K. Marx and F. Engels*, Vol. 2. Beijing: People's Publishers. 217 (Chinese Edition).

McCarthy, J. and P. Hayes. 1969. Some philosophical problems from the standpoint of artificial intelligence. [D. Michie, ed., *Machine Intelligence 4*. New York: American Elsevier.

Neves, D. M. 1978. A computer program that learns algebraic procedures by examining examples and by working test problems in a textbook. *Proceedings of the Second Annual Conference of the Canadian Society for Computational Studies of Intelligence*. Toronto, Canada.

Pavlov, I. P. 1951(a). Attempts in the physiological interpretation of hysterical symptoms. In *Collected Works of I. P. Pavlov*, Vol. III(2). Moscow: USSR Academy of Sciences Press, 195-218 (In Russion).

Pavlov, I. P. 1951(b). The task and the arrangement of a laboratory for the study of the normal activity of the highest parts of the central nervous system in the higher animals. In *Collected Works of I. P. Pavlov*, Vol. III(1). Moscow: USSR Academy of Sciences Press. 127-146 (In Russian).

Rosenzweig, M. R. 1985. Neuronal plasticity related to cognition. In *Psychophysiological Approaches to Human Information Processing*, (Edited by F. Klix, R. Naatanen, and K. Zimmer). Amsterdam: North-Holland. 31-45.

Rummelhart, D. E., and J. L. McClelland. 1986. *Parallel Distributed Processing, Volume 1, Foundations*. Cambridge, Mass.: MIT Press.

Selfridge, O. G. 1959. Pandemonium: A paradigm for learning. In *Proceedings of a Symposium on the Mechanization of Thought Processes*. London: H. M. Stationery Office.

Simon, H. A. 1986. The information processing explanation of Gestalt phenomena. *Computers in Human Behavior*, 2:241-255.

Simon, H. A., and M. Barenfeld. 1969. Information processing analysis of perceptual processes in problem solving. *Psychological Review*, 76: 473-483.

Simon, D. P., and H. A. Simon. 1978. Individual differences in solving physics problems. [R. S. Siegler, ed., *Children's Thinking: What Develops?*. Hillsdale, N.J.: Erlbaum Associates.

Sokolov, E. N. 1963. *Perception and the Conditioned Reflex*. New York: Pergamon Press.

Weyhrauch, R. W. 1980. Prolegomena to a theory of mechanized formal reasoning. *Artificial Intelligence* 13:133-170.

Zadeh, L. (ed.). 1975. *Fuzzy Sets and Their Applications*. N. Y.: Academic Press.